Joint Scaling Theory of Human Dynamics and Network Science

Chaoming Song,[1,2] Dashun Wang,[1,2] and Albert-László Barabási[1,2,3]

[1] *Center for Complex Network Research, Department of Physics, Biology and Computer Science, Northeastern University, Boston, Massachusetts 02115, USA*

[2] *Center for Cancer Systems Biology, Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA*

[3] *Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA*

Abstract

The increasing availability of large-scale data on human behavior has catalyzed simultaneous advances in network theory, capturing the scaling properties of the interactions between a large number of individuals, and human dynamics, quantifying the temporal characteristics of human activity patterns. These two areas remain disjoint, however, traditionally each pursuing as a separate modeling framework. Here we establish the first formal link between these two areas by showing that the exponents characterizing the degree and link weight distribution in social networks can be expressed in terms of the dynamical exponents characterizing human activity patterns. We test the validity of these theoretical predictions on datasets capturing various facets of human interactions, from mobile calls to tweets. We find evidence of a universal measure, that links networks and human dynamics, but whose value is independent of the means of communication, capturing a fundamental property of human activity.

Fueled by data collected by a wide range of high-throughput tools and technologies, the study of complex systems is currently reshaping a number of research fields, from cell biology to computer science. Nowhere are these advances more apparent than in the study of human dynamics and social media. Indeed, the unparalleled use of email, mobile devices and social networking have provided researchers access to massive amounts of data on the real time activity patterns of millions of individuals, simultaneously fueling advances in two research areas, network science [1–4] and human dynamics [5–8]. Network science focuses on the structure and dynamics of complex networks that capture the totality of interactions between individuals, having led to the discovery of a series of generic properties of real networks, from the fat tailed nature of the degree distribution [9, 10] to predictable patterns characterizing the weights or link strengths [11–13]. Human dynamics in contrast focuses on the temporal aspects of individual interaction patterns, offering evidence that the interevent times between consecutive events initiated by an individual follow a fat tailed distribution [5, 6, 14–21], representing a significant deviation from a Poisson process predicted by random communications. As network theory [1, 3, 4, 9] and human dynamics [5, 14] have developed in parallel, being pursued as separate lines of inquiry, we lack relationships between the quantities explored by them, despite the fact that they often study the same systems and datasets. Here we derive a series of generic laws that link the quantities characterizing social networks and human dynamics, showing that the widely studied properties uncovered independently in the two areas represent two facets of the same underlying phenomena.

To demonstrate the practical relevance of our results, we compiled four independent datasets that together capture most aspects of digital communication that humans are involved in lately (SI Section 1): 1) *Mobile phone data*, that summarizes the communication patterns of about 4 million anonymized European mobile users during a year period, providing access to over 1.2 billion events, representing information on who talks with whom and the timing of each call [22]; 2) *E-mail traffic* within a university, that collects over two million email messages sent during an 83 day period exchanged by around 3,000 users [14, 23]; 3) *Twitter data*, that records the tweets of about 0.7 million users, containing over 8 million messages collected between Aug 2009 and Mar 2010 [24–26]. 4) *Online Messages*, that records more than 500,000 messages sent by approximately 30,000 active users of a Swedish dating site over 492 days [6, 27].

Two widely studied quantities characterize the underlying social networks:

Degree distribution: The degree $k_i(t_1, t_2)$ of an individual i represents the total number of individuals he/she contacted within the $[t_1, t_2]$ time interval, including both acquaintancy and strong ties [28, 29]. The degree distribution $P_k(k) \equiv N^{-1} \sum_{i=1}^{N} \delta(k - k_i)$ of each studied systems can be approximated with a power law [9, 10, 22] (Fig. 1b).

$$P_k(k) \sim k^{-\gamma_k}, \tag{1}$$

where the degree exponent varies between $\gamma_k = 1.0$ for Twitter and $\gamma_k = 4.8$ for mobile phones (Table I). The measurements indicate that for Twitter, email, and online messages γ_k is independent of time, but for mobile phones decreases from $\gamma_k = 4.19$ to $\gamma_k = 3.20$ during a year (P_k for different time intervals is shown in SI Section 7).

Weight distribution: Denoting with $w_{i \to j}$ (weight) the number of contacts between two nodes [11–13], we measure the weight distribution $P_w(w) \sim \sum_{i,j} \delta(w - w_{i \to j})$ for different dataset (Fig. 1c), finding that the weight distribution can be approximated with (Fig. 1c)

$$P_w(w) \sim w^{-\gamma_w}, \tag{2}$$

where the weight exponent varies between $\gamma_w = 1.51$ for mobile phones and $\gamma_w = 1.9$ for emails (Table I).

To explore the dynamics of human activity, we focus on two frequently measured quantities [5, 14, 30]:

Activity distribution: Denoting with $C_i(t_1, t_2)$ the activity, representing the total number of communications initiated by individual i within a $[t_1, t_2]$ time interval, we find that the activity distribution $P_C(C) \equiv N^{-1} \sum_{i=1}^{N} \delta(C - C_i)$ is fat tailed, following (Fig. 1d)

$$P_C(C) \sim C^{-(1+\beta_C)}, \tag{3}$$

where β_C ranges between 0.1 (Twitter) to 3.38 (mobile phones) (Fig. 1d and Table I, P_C for different time intervals is shown in SI Section 7).

Interevent time distribution: A key property of human dynamics is the non-Poissonian nature of the interevent time Δt between consecutive communication patterns [14–19, 31]. Previous studies have found that $P_{\Delta t}(\Delta t) \sim \Delta t^{-\beta_0} \exp(-\Delta t / \Delta t_c)$, with $\beta_0 \simeq 1$ (SI Section 3.1 and Refs. [14–19, 31]). As $P_{\Delta t}(\Delta t)$ characterizes the communications between *all* friends, here we define a link-specific interevent time $\tau_{i \to j}$ as the total number of communication events initiated by user i between two consecutive communications from i to j [32]. For

example, $\tau_{A\to C} = 3, 4, 5$ in Fig. 1a. We measure the probability density function $P_{\tau,i}(\tau)$ for each individuals, finding they that all follow broad distributions (see SI Section 6). In Fig. 1e, we plot $P_\tau(\tau) \equiv N^{-1}\sum_i C_i^{-1} P_{\tau,i}(\tau/C_i)$ averaged over users with $C_i > 100$ (see Fig. S6 for $P_{\tau,i}$ for different C_i activity groups), finding that it is also fat-tailed, well approximated by (Fig. 1e)

$$P_\tau(\tau) \sim \tau^{-(1+\beta_\tau)}, \tag{4}$$

where β_τ characterizes the inhomogeneity of the communication pattern for a pair of users. Queuing models predict $\beta_\tau = \beta_0 = 0$ (fixed queue length) or 0.5 (variable queue length) [14–19, 31]. We find, however, that the empirically observed β_τ varies between 0.2 (online messages) and 0.53 (mobile phones) (Table I), differences not accounted for by the existing theoretical approaches.

In summary, the underlying social network is characterized by $P_k(k)$ and $P_w(w)$, while the communication dynamics by $P_\tau(\tau)$ and $P_C(C)$, each with its system dependent form. These two classes of phenomena, and the associated distributions, are treated independently in the literature [1, 3–19]. Next we derive a set of relationships between the two phenomena and help uncover a universal quantity that appears to be independent of the means of communication used by an individual.

While one expects that the more active is an individual (high C_i), the more friends he/she has (high k_i), as shown in Fig. 1b,e and Table I, the distributions $P_k(k)$ and $P_C(C)$ are not equivalent. To understand the relationship between k_i and C_i, we measured for each individual how their degree (k_i) grows with the number of communication events (C_i) they participated in. We find that the individual degree k_i can be approximated with (Fig. 2a)

$$k_i(t_1, t_2) \sim C_i(t_1, t_2)^{\alpha_i}, \tag{5}$$

where the exponent α_i, which characterizes the individual's affinity to translate its level of activity into new contacts, varies from individual to individual. For each user $\alpha_i < 1$, indicating that the degree grows sub-linearly with the activity C_i, a phenomenon known as Heaps' law in linguistics [33]. This means that increasing the number of calls has diminishing impact on the growth in the number of friends. While the temporal patterns of both k_i and C_i might be affected by environmental factors and circadian rhythms, we find that Eq. (5) is independent of the observational time frame.

The fact that the exponent α_i varies from individual to individual indicates that users with similar activity levels acquire degrees at different rates (Fig. 2a). Therefore, α_i characterizes an individual's ability to add friends given his/her activity level C_i, prompting us to call α_i *sociability*. To investigate the demographic variation of sociability, in Fig. 2b we show the sociability distribution for all four datasets, finding that $P_\alpha(\alpha) \equiv N^{-1} \sum_i \delta(\alpha - \alpha_i)$ is bounded between 0 and 1 and decays rapidly on both sides of the peak. We also find that α_i is largely independent of C_i, as indicated by the conditional probability $P_\alpha(\alpha|C)$, that overlaps for users with different activity C (Fig. 2b, inset). Somewhat surprisingly, this indicates that sociability, i.e., the ability to establish new contacts, is independent of the individual's activity level, representing instead an intrinsic property of an individual. Figure 2b shows $P_\alpha(\alpha)$ for all datasets, indicating that each communication system is characterized by its own distinct $P_\alpha(\alpha)$ and average sociability $\overline{\alpha}$ (see Table I & inset of Fig. 2c).

Heaps' law (5) allows us to link a node's degree, a network measure, to the bursty patterns of human activity. Indeed, let us denote with $\Pi_i(t_1, t_2)$ the probability that individual i contacts a *new* friend j, i.e. someone that i did not contact in the previous $[t_1, t_2]$ time frame. This requires that the waiting time $\tau_{i \to j}$, that characterizes the communication between i and j, be greater than $C_i(t_1, t_2)$. The probability that the interevent time exceeds $C_i(t_1, t_2)$ is $\Pi_i(t_1, t_2) = \int_{C_i}^{\infty} P_{\tau,i}(\tau) d\tau \sim C_i^{-\beta_{\tau,i}}$. Yet, (5) indicates that $\Pi_i = dk_i/dC_i \sim C_i^{\alpha_i - 1}$. Comparing these two equations, we obtain the scaling relationship

$$\alpha_i + \beta_{\tau,i} = 1. \tag{6}$$

Equation (6) indicates that the bursty nature of human activity patterns ($\beta_{\tau,i}$) determines the growth of individual degree in the social network via $k_i \sim C_i^{1-\beta_{\tau,i}}$, predicting the observed sublinear growth. Therefore, Eq. (6) represents the first bridge between human dynamics (e.g. $\tau_{i \to j}$ and $\beta_{\tau,i}$) and the structure of the underlying social network (e.g. k_i and α_i). Equation (6) offers a rather strong prediction, indicating that a simple scaling relationship holds for each individual, despite the individual variability in α_i and $\beta_{\tau,i}$. For validation we need to determine α_i and $\beta_{\tau,i}$ for each individual, which requires sufficient individual statistics. However, as sociability α_i is largely independent of individual activity level (inset of Fig. 2b), we can focus on active users, for which we have sufficient data, without introducing a selection bias. In Fig. 2c, we measure the sociability α_i and the waiting time exponent $\beta_{\tau,i}$ independently for users with $C_i > 1000$ and $k_i > 100$ in the mobile phone dataset (see

SI Section 6), finding that while α_i and $\beta_{\tau,i}$ vary considerably between users, $P(\alpha_i + \beta_{\tau,i})$ is narrowly peaked around 1, offering strong support for Eq. (6). As shown in Table I, the prediction (6) is consistent not only with the exponents measured in each dataset, but also with the queuing models [14, 15, 31] (for sake of simplicity, the average $\overline{\alpha}$ and $\overline{\beta_\tau}$ are reported). Perhaps most surprisingly, we find that when rescaled with the average $\overline{\beta_\tau}$, $P_{\beta_\tau}(\beta_\tau)$ for the different datasets collapse into a single curve (Fig. 2d)

$$P_{\beta_\tau}(\beta_\tau) = (1/\overline{\beta_\tau})F(\beta_\tau/\overline{\beta_\tau}), \tag{7}$$

suggesting that the distribution $P_{\beta_\tau}(\beta_\tau)$ of the bursty exponent β_τ captures an inherent property of the population, independent of the means of communication. This data collapse is quite remarkable, given the difference in the nature of the data (calls, emails, tweets, and online messages), timeframes, countries and demographics (phone: about 25% of an European country's population [22]; emails: university employees from a different European country [14, 23]; Twitter: mainly US [24, 25]; Online Messages: Swedish teenagers [6, 27]). Figure 2d suggests an exponential growth of $F(x)$ for small x, i.e., $F(x) \sim \exp(\sigma x)$, where $\sigma \approx 6.6$ appears to be the same for all datasets (Table I), a parameter that will play an important role below.

The scaling law (5), together with the sociability distribution $P_\alpha(\alpha)$ allows us to derive an another relationship between social networks and human dynamics. Indeed, the statistical independence between α and C implies

$$P_k(k) = \int \delta(k - C^\alpha) P_\alpha(\alpha) P_C(C) d\alpha dC, \tag{8}$$

indicating that the fat tailed nature of the degree distribution is rooted in the population heterogeneity in terms of sociability α_i and activity C_i. Note that this relationship is independent of the particular form of $P_k(k)$ and $P_C(C)$, being equally valid if they follow power laws, stretched exponentials or log-normal distributions. We compared the empirically measured P_k with the prediction (8) for all datasets, obtaining excellent agreement (Fig. 3a-d). Therefore, Eq. (8) is the second bridge between human dynamics and the network structure, capturing the competition between two phenomena:

Activity Driven: If $P_k(k)$ is dominated by differences in the users' activity level (the activity distribution $P_C(C)$), we can ignore the variations in P_α, replacing individual sociability (α_i) with $\overline{\alpha}$, finding

$$P_k(k) \sim k^{1/\overline{\alpha}-1} P_C(k^{1/\overline{\alpha}}). \tag{9}$$

This limit correctly describes email, twitter, and online messages (Fig. 3a-c).

Sociability Driven: If $P_\alpha(\alpha)$ dominates, the individuals' activity level (C_i) can be approximated with their mean $\overline{C}$, and Eq. (7) predicts that the sociability distribution has an exponential tail $P_\alpha(\alpha) \sim \exp(-\alpha\sigma/\bar{\beta}_\tau)$) (shaded area in Fig. 2b) that dominates the scaling of (8), obtaining

$$P_k(k) \sim k^{-\left(1+\sigma/(\overline{\beta}_\tau \ln \overline{C})\right)}, \tag{10}$$

indicating P_k has a power law tail, whose exponent γ_k is determined by variability in sociability, captured by the parameter σ. More interestingly, it predicts that γ_k decreases with the average activity level $\overline{C}$, leading to a scaling exponent that depends on an extensive quantity, not observed before in network science. Indeed, as $\overline{C}$ increases with the observation time (Fig. S2), (10) predicts a time-dependent γ_k, driven by changes in $\overline{C}$. Figure 3e-f show that despite the temporal stationarity of individual activity (Fig. S8A) for mobile communications, γ_k decreases with $\overline{C}$ for different time interval $[t_1, t_2]$, indicating that the degree heterogeneity of mobile phone users is indeed driven by variability in their sociability.

Combining (3) with these two classes, we predict the degree exponent in (1), as (SI Section 4)

$$\gamma_k = 1 + \min\left[\frac{\beta_C}{1-\overline{\beta}_\tau}, \frac{\sigma}{\overline{\beta}_\tau \ln \overline{C}}\right]. \tag{11}$$

In Table I we report the γ_C and γ_k of the power law model for all datasets. Yet, Eqs. (8-10) are not limited to power laws; other fat tailed models for P_C such as lognormal or stretched exponential can also be exploited, as discussed in SI Section 4. The fundamental relationship (8) and the distinction between the two classes is therefore independent of particular models (and fits) for P_C.

To derive the network's weight distribution $P_w(w)$ we note that for each individual i, $\sum_j w_{i\to j} = C_i$, where $w_{i\to j}$ denotes the total number of messages/calls from i to j. We denote with $p_r \equiv p_{i\to j} \equiv w_{i\to j}/C_i$ the probability that user i communicates with user j, and r is the rank of $p_{i\to j}$ across all friends j of user i. We find that p_r is well approximated by Zipf's law $p_r \sim r^{-\zeta_i}$ (Fig. 4a) [34], a direct consequence of the fat tailed nature of $P_w(w)$ [11–13]. That is, an individual communicates most of the time with only a few individuals and it interacts with the rest of its contacts with diminished frequencies. Intuitively, one would assume ζ is the same for individuals with the same activity C. Yet, we find that for three randomly selected users, each with the same activity $C_i = 400$, p_r has different

ζ_i exponents (Fig. 4a). However, for users with different activities but the same sociability α, the curves are indistinguishable (Fig. 4a), hinting the existence of a link between ζ_i and α_i. This relationship can be derived by focusing on an individual's least preferred contact. Intuitively, there are only a few communications ($O(1)$) between the individual i and his/her least preferred contact, independent of the activity level C_i. Therefore, given k_i, the total number of contacts of individual i is $C_i p_{k_i} = C_i k_i^{-\zeta_i} = C_i^{1-\alpha_i\zeta_i} = O(1)$, obtaining [35, 36],

$$\alpha_i \zeta_i = 1. \tag{12}$$

The validity of (12) is confirmed in Fig. 4b, where we show that $p_r(r^{1/\alpha})$ collapses for users with different α_i and the curve has the slope -1 for the top ranked contacts, as predicted by (12). Equation (12) allows us to derive the weight distribution $P_w(w)$. The weight distribution $P_w(w)$ is averaged over populations, as

$$P_w(w) = \int \sum_{r=1}^{C^\alpha} \delta(w - A(C,\alpha) C r^{-1/\alpha}) P_C(C) P_\alpha(\alpha) dC d\alpha, \tag{13}$$

where the normalization factor $A(C,\alpha) \equiv \mathcal{A} \sum_{r=1}^{C^\alpha} r^{-1/\alpha}$ with a system-dependent constant $\mathcal{A}$ corresponding to the average weight. Figure 4c-f confirms the validity of Eq. (13) for all datasets. The fact that Zipf's law is equivalent with $P_{w,i}(w) \sim w^{-(1+1/\zeta_i)} = w^{-(1+\alpha_i)}$, where $P_{w,i}(w)$ represent the weight distribution of individual i, leads to a first order approximation of Eq. (13) as $P_w(w) \sim w^{-\gamma_w}$ with cutoff, where the exponent $\gamma_w = 1 + \overline{\alpha} = 2 - \overline{\beta}_\tau$ up to the leading order. Combining this with (6), we find

$$\gamma_w = 2 - \overline{\beta}_\tau. \tag{14}$$

The prediction (14), supported by the empirical data in Table I, represents the third link we uncover between a temporal property of human dynamics (β_τ) and a structural property of social networks (γ_w).

In summary, Eqs. (4), (6), (8), (12) and (13) offer five direct links between human dynamics and the architecture of social networks, showing that the degree distribution (P_k) and the tie strength distribution (P_w) can be expressed in terms of the dynamical exponents characterizing the temporal patterns in human activity, like burstiness ($P_{\tau,i}$) and the activity level (P_C). The relationship between these two classes of exponents is mediated by the parameter σ, which we find to be independent of the communication technology, hence capturing an inherent property of human activity. These relationships bring an unexpected

order to the zoo of exponents reported in Table I, showing that they represent different facets of a deeper underlying reality.

A better understanding of the origin of the exponents reported in Table I requires mechanistic models, tailored to the specific communication phenomena. Yet, the relationships (4), (6), (8), (12), and (13) derived here are independent of the system's details or the specific communication mechanism, thus all future models that aim to account for human dynamics and social networks in a specific system must obey them. Hence, the theory developed here represents the skeleton of a joint mathematical theory of human dynamics and social networks. As our understanding of human dynamics deepens with the emergence of new and increasingly detailed data on both human activity patterns and social networks, such fundamental relationships are expected to have increasing value, helping us anchor future models and offer a springboard towards a deeper mechanistic understanding of big data, the often noisy, incomplete, but massive datasets that trail human behavior.

The authors wish to thank A. Mislove for providing the Twitter dataset and P. Holme for providing the Online Messages dataset. This work is supported by the NSF (IIS-0513650); ONR (N000141010968); DTRA (HDTRA1-08-1-0027); DARPA (11645021) and the Network Science Collaborative Technology Alliance sponsored by ARL (W911NF-09-2-0053).

[1] G. Caldarelli, *Scale-Free Networks* (Oxford Finance, 2007).

[2] R. Cohen and S. Havlin, *Complex networks: structure, robustness and function* (Cambridge Univ Pr, 2010).

[3] M. Newman, *Networks: an introduction* (Oxford University Press, 2010).

[4] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks* (Oxford University Press, 2003).

[5] C. Castellano, S. Fortunato, and V. Loreto, Reviews of Modern Physics **81**, 591 (2009).

[6] D. Rybski, S. Buldyrev, S. Havlin, F. Liljeros, and H. Makse, Proceedings of the National Academy of Sciences **106**, 12640 (2009).

[7] D. Lazer *et al.*, Science **323**, 721 (2009).

[8] D. Brockmann, L. Hufnagel, and T. Geisel, Nature **439**, 462 (2006).

[9] R. Albert and A.-L. Barabási, Reviews of Modern Physics **74**, 47 (2002).

[10] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[11] K.-I. Goh, B. Kahng, and D. Kim, Physical Review Letters **87**, 278701 (2001).

[12] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, Proceedings of the National Academy of Sciences **101**, 3747 (2004).

[13] A. Barrat, M. Barthélemy, and A. Vespignani, Physical Review Letters **92**, 228701 (2004).

[14] A.-L. Barabási, Nature **435**, 207 (2005).

[15] J. Oliveira and A.-L. Barabási, Nature **437**, 1251 (2005).

[16] A. Vázquez, Physical Review Letters **95**, 248701 (2005).

[17] A. Vázquez, J. Oliveira, Z. Dezsö, K.-I. Goh, I. Kondor, and A.-L. Barabási, Physical Review E **73**, 036127 (2006).

[18] A. Gabrielli and G. Caldarelli, Physical Review Letters **98**, 208701 (2007).

[19] A. Gabrielli and G. Caldarelli, Physical Review E **79**, 041133 (2009).

[20] R. Malmgren, D. Stouffer, A. Campanharo, and L. Amaral, Science **325**, 1696 (2009).

[21] J. L. Iribarren and E. Moro, Social Networks **33**, 134 (2011).

[22] J. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, Proceedings of the National Academy of Sciences **104**, 7332 (2007).

[23] J. Eckmann, E. Moses, and D. Sergi, Proceedings of the National Academy of Sciences **101**, 14333 (2004).

[24] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, in *4th International AAAI Conference on Weblogs and Social Media (ICWSM)* (2010).

[25] A. Mislove, S. Lehmann, Y. Ahn, J. Onnela, and J. Rosenquist, in *5th International AAAI Conference on Weblogs and Social Media (ICWSM)* (2011).

[26] S. Golder and M. Macy, Science **333**, 1878 (2011).

[27] P. Holme, C. R. Edling, and F. Liljeros, Social Networks **26**, 155 (2004).

[28] R. Dunbar, Journal of Human Evolution **22**, 469 (1992).

[29] B. Gonçalves, N. Perra, and A. Vespignani, PLoS One **6**, e22656 (2011).

[30] J. L. Iribarren and E. Moro, Phys. Rev. Lett. **103**, 038702 (2009).

[31] A. Cobham, Journal of the Operations Research Society of America **2**, 70 (1954).

[32] M. Karsai, K. Kaski, A. Barabási, and J. Kertész, Scientific Reports **2**, 397 (2012).

[33] H. Heaps, *Information retrieval: Computational and theoretical aspects* (Academic Press, Inc. Orlando, FL, USA, 1978).

[34] B. Corominas-Murtra, J. Fortuny, and R. V. Solé, Physical Review E **83**, 32767 (2011).

[35] R. Baeza-Yates and G. Navarro, Journal of the American Society for Information Science **51**, 69 (2000).

[36] L. Lu, Z. Zhang, and T. Zhou, PloS one **5**, e14139 (2010).

FIGURES

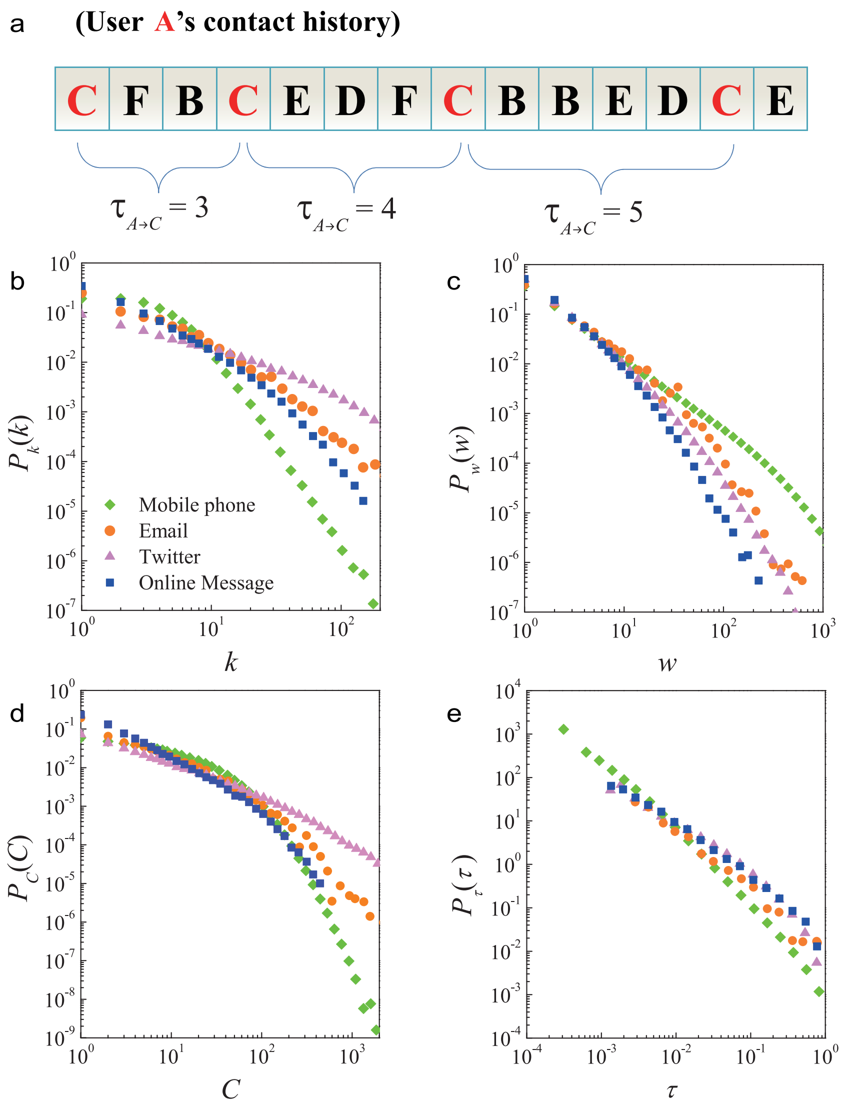

FIG. 1. **Basic measures characterizing networks and human dynamics.** (a) The definition of $\tau_{A\rightarrow C}$, the interevent time captures communication intervals between two individuals, A and C. Note that $\tau_{A\rightarrow C}$ measures time in terms of the number of events, a feature that corrects for daily fluctuations in the communication volume, but has the same asymptotic scaling as the real interevent time [23]. (b) Degree distribution $P_k(k)$, and (c) link weight distribution $P_w(w)$ for each of the four studied datasets. (d) Activity distribution $P_C(C)$. (e) The distribution of the number events between consecutive communications with the same individual, $P_\tau(\tau)$, where τ is normalized by each individual's activity level C.

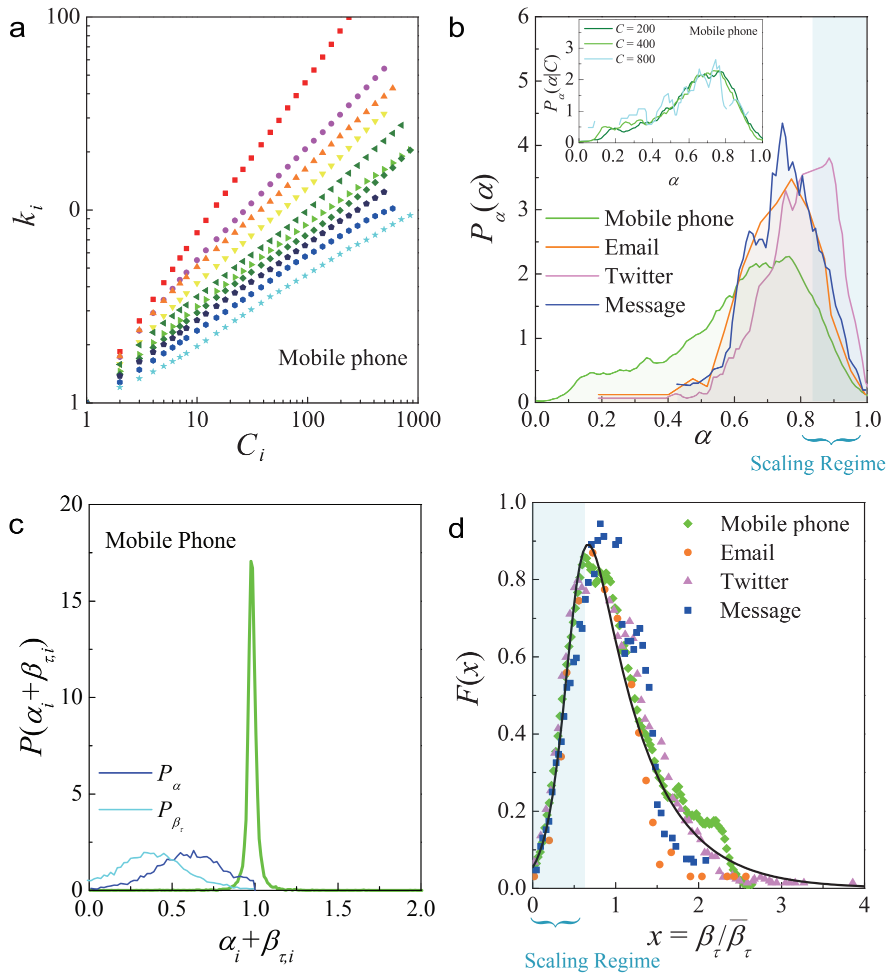

FIG. 2. **Measuring user sociability.** (a) The growth in degree $k_i(t_1, t_2)$ for ten mobile phone users in function of the same user's activity $C_i(t_1, t_2)$, where each dot corresponds (C_i, k_i) for one time frame $[t_1, t_2]$. Similar curves are observed for the other datasets (see Fig. S3). (b) The sociability distribution, $P_\alpha(\alpha)$, for the three studied datasets, where the shaded region highlights the tail of $P_\alpha(\alpha)$. Inset: conditional probability distribution $P_\alpha(\alpha|C)$ for mobile phone users with activity $C = 200$, 300 and 800, respectively. (c) $P(\alpha_i + \beta_{\tau,i})$ (green curve) measured for the mobile phone dataset for users with $C_i > 1000$ and $k_i > 100$, supporting the validity of Eq. (6). The purple and blue curves correspond P_α and P_{β_τ} for the same user group, respectively. (d) The collapse of $P_{\beta_\tau}(\beta_\tau)$ distributions after rescaling $P_{\beta_\tau}(\beta_\tau)$ with average $\overline{\beta_\tau}$ for each datasets. The black line represents a Burr type II distribution, $F(x) \propto \exp(\sigma x)/\left(1 + s\exp(\kappa x)\right)$ with $\sigma = 6.6$, capturing the exponential growth $F(x) \sim \exp(6.6x)$ for small β_τ.

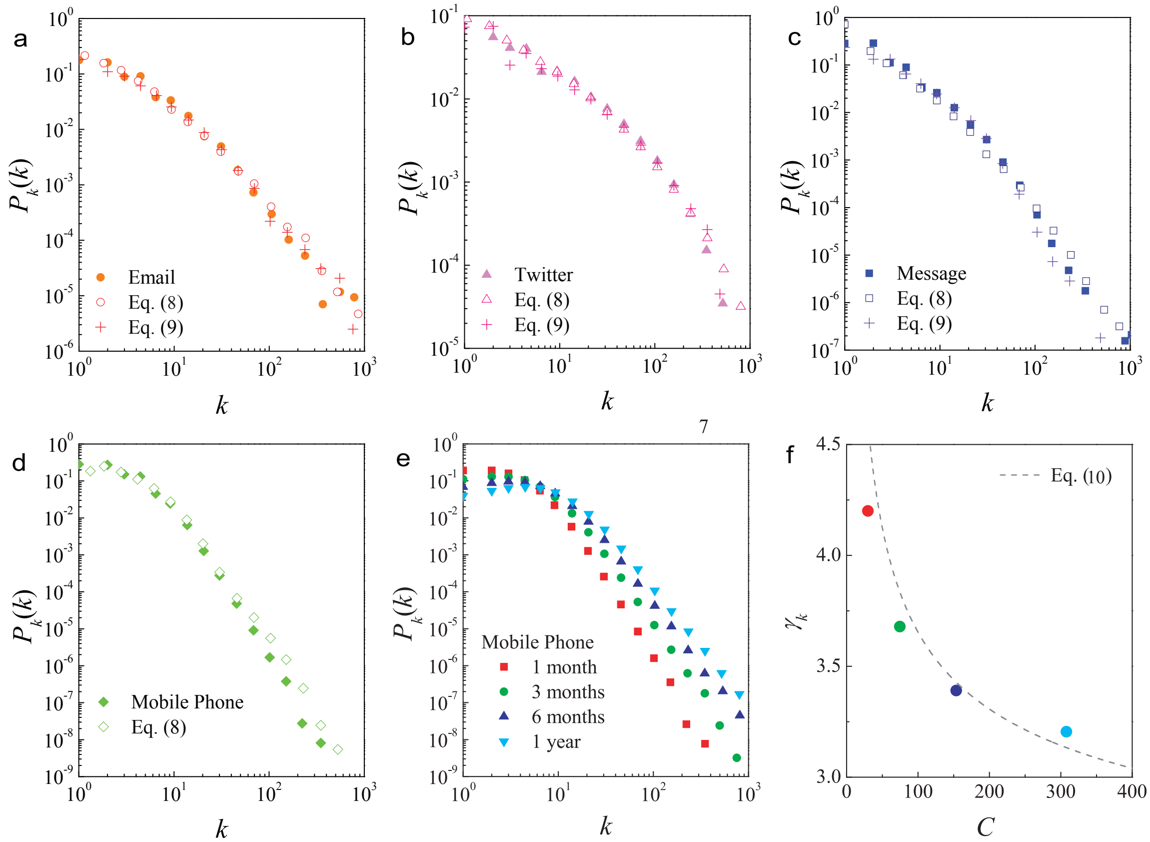

FIG. 3. **Predicting the Degree Distribution** The measured degree distribution $P_k(k)$ (solid), compared to the predictions of Eqs. (8) (open) and (9) (cross) for (a) Email (b) Twitter (c) Online Message and (d) Mobile Phone datasets, respectively, showing that Eq. (8) is consistent with empirical observation. For emails we adjusted (8) to allow for multiple recipients (see SI Section 2). The validation of Eq. (9) for Email, Twitter and Online Message datasets also indicates these systems are *activity driven*. (e) $P_k(k)$ for mobile phone dataset, revealing power law tails for different time frames $\Delta T \equiv t_2 - t_1$, from 1 month to 1 year (see SI Section 7 for all datasets). (f) The degree exponent γ_k deceases with average activity $\overline{C}$ as predicted by (10), indicating that mobile phone communication is *sociability driven*.

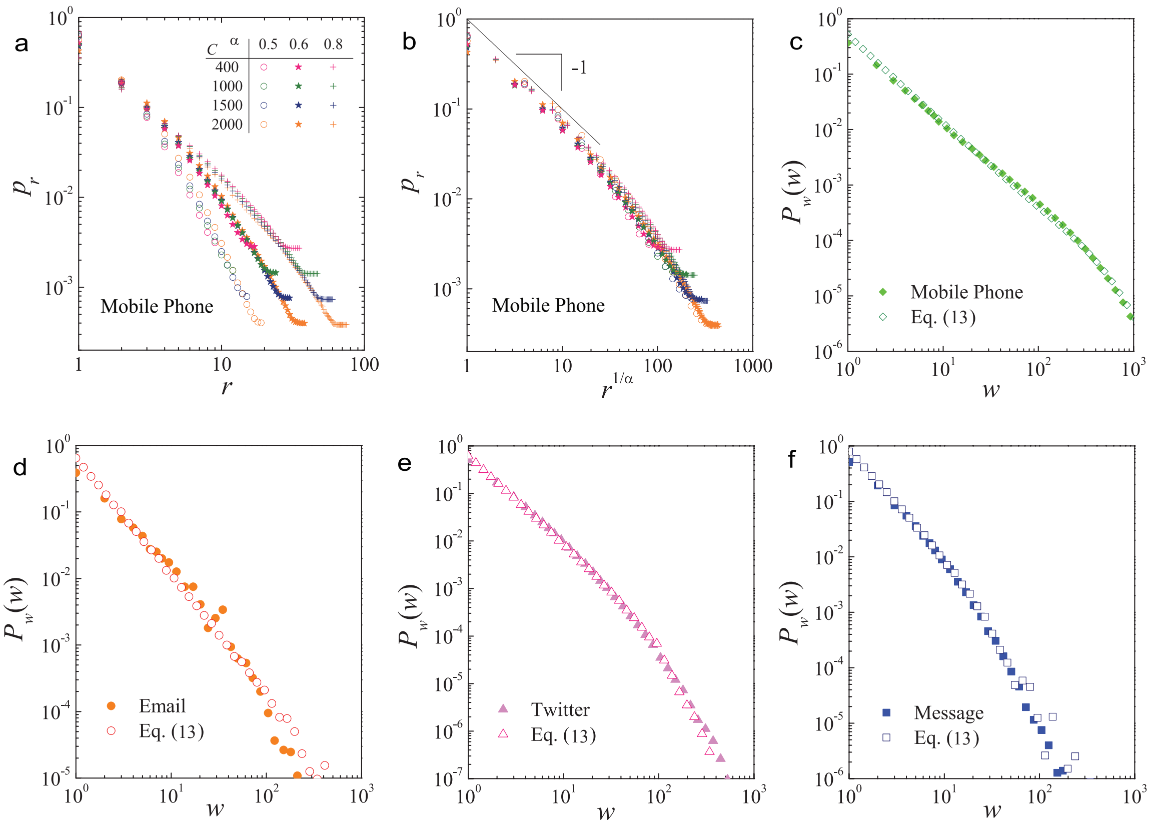

FIG. 4. **Quantifying the tie strength distribution.** (a) Zipf's plot showing the communication frequency $p_{r,i}$ for a user i with the user's r-th most contacted friend for the mobile phone data. The different colors and symbols represent different activities and sociabilities, respectively, indicating that the Zipf's exponent ζ_i depends only on the sociability α_i. (b) The plot of p_r versus $r^{1/\alpha}$ showing collapses over different sociability groups, as predicted by $\alpha_i \zeta_i = 1$, derived in (12). Similar plots are observed for the other datasets (see SI Section 3.3). (c,d,e,f) The degree distribution $P_w(w)$ from empirically measurements (solid), comparing to the predictions of Eq. (13) for (c) Mobile Phone (d) Email (e) Twitter and (f) Online Message datasets, respectively, showing that Eq. (13) is consistent with the empirical observation.

TABLES

TABLE I. **Quantify networks and human dynamics.** The scaling exponents characterizing the networks and human dynamics in the four studied datasets, as well as the most studied human dynamics models. The reported $\overline{\alpha}$ and $\overline{\beta_\tau}$ represent average values over the population for empirical data, where $\overline{\beta_\tau}$ is measured from $P_\tau(\tau) \sim \tau^{-(1+\overline{\beta_\tau})}$ as a first order approximation. The error of $\overline{\beta_\tau}$ and β_C are derived from the error of $1+\overline{\beta_\tau}$ and $1+\beta_C$, respectively. Note that the small error bars of exponents are due to the large population size. See SI Section 6 for justification of the goodness of fit.

	Mobile phone	Email	Twitter	Message	Queueing Models	
					Fixed Length [31]	Variable Length [14]
γ_k	$4.19_{\pm 0.01} \div 3.205_{\pm 0.007}$	$2.27_{\pm 0.01}$	$1.241_{\pm 0.001}$	$1.624_{\pm 0.003}$	–	–
γ_w	$1.51335_{\pm 0.00006}$	$1.637_{\pm 0.003}$	$1.8483_{\pm 0.0006}$	$1.930_{\pm 0.002}$	–	–
$\overline{\beta_\tau}$	$0.53823_{\pm 0.00001}$	$0.431_{\pm 0.002}$	$0.3162_{\pm 0.0001}$	$0.360_{\pm 0.002}$	0	0.5
β_C	$3.39_{\pm 0.01}$	$0.82_{\pm 0.01}$	$0.147_{\pm 0.001}$	$0.430_{\pm 0.002}$	–	–
$\overline{\alpha}$	$0.58_{\pm 0.01}$	$0.68_{\pm 0.02}$	$0.78_{\pm 0.01}$	$0.70_{\pm 0.01}$	1.0	0.5
σ	$6.6_{\pm 0.1}$	$6.8_{\pm 0.2}$	$6.6_{\pm 0.1}$	$6.6_{\pm 0.1}$	–	–
$\ln \overline{C}$	$3.4 \div 5.9$	4.8	5.4	3.0	–	–